Reflection of @SCAD.edu - A virtual mirror on E-mail communication at an art and design university

Marc Egger\textsuperscript{a}, Hauke Fuehres\textsuperscript{b,*}

\textsuperscript{a}University of Cologne, Albertus-Magnus-Platz, 50923 Cologne, Germany
\textsuperscript{b}University of Applied Sciences Northwestern Switzerland, Steinackerstrasse 5, 5210 Windisch, Switzerland

Abstract

This article describes an information system visualizing the E-mail communication of students and staff at the Savannah College of Art and Design (SCAD). In contrast to other visualization tools the described software is to be exhibited in a public space without the possibility of sophisticated user interaction. In such a setting addressing privacy issues is a key factor to the acceptance of the software. Besides visualizing the E-mail communication network the system also takes a look at the social media service Twitter to gain an impression on the discussed topic in the gathered E-mails.

Keywords: Social Network Analysis (SNA); E-mail network; Visualization
1. Introduction

E-mail has become a key medium in communication frequently used at universities. This article describes an information system exposing e-mails as anonymized communication entities, and making visible the arising social network of students and staff at the Savannah College of Art and Design (SCAD). The introduced system offers a dynamic view on the communication network of the participants. It offers a live-view on the evolving communication structure. In order to create self-awareness on their own social network, participants can obtain a live-view of their very own communication behavior.

The introduced visualization places particular demands on the data gathering process needed to collect the e-mails: Showing the e-mail communication at the moment of its occurrence requires continuous real-time collecting and processing of the messages sent.

Besides visualizing the social network constituted by the messages sent at SCAD the presented system also takes a look at the topics the participants are discussing. The discussed topics are reflected by the social media service Twitter, allowing to see what the Twitter users think and feel about topics mentioned in the messages collected. Extracting those topics to be mirrored in Twitter is detached from particular e-mail addresses to maintain the privacy of each message sent. In order to pique the curiosity of users of the e-mail visualization and prospective participants the sentiment extracted from the social media service is ciphered in QR codes, which can be deciphered with cell phones.

It is aimed to visualize the communication without the need of further interaction.

Crucial to the acceptance of the described system is the respect for privacy of the participants. Users have to opt-in to use the system. Senders and receivers of the emails are anonymized. Topics extracted from the e-mail subjects cannot be linked to particular communication.

A part of the field of Information Visualization deals with the visualization of social networks and the processing of those networks to allow gaining new insights into their structure. These networks can be of different nature. E-mail communication networks were visualized (Gloor et al. 2003), as well as the link structure of online social networking sites like Friendster.com (Heer & Boyd 2005).

2. Setting (University E-mails)

The introduced system aims to visualize the social network of students and staff at SCAD and its evolution over time. People can individually decide to take part in the visualization of their university’s E-mail network. In contrast to other E-mail visualization toolkits, the SCAD visualization is not accessing a single, downloaded E-mail box or pooled E-mail boxes but needs to obtain data from several data sources in a near real-time fashion (Gloor & Zhao 2004).

3. Privacy – a main driver of the system

The developed software is built to be exhibited in a public space. A main concern during the conception of the system was to respect the privacy of the participants when other people observe the exposed communication patterns.

Privacy issues addressed by our system can be divided in two parts:
1. Ensuring the privacy of the participants towards the users of the systems. Viewers of the visualization should not identify participating people nor should the content of the participants’ E-mails be exposed.
2. The gathered data from the participants’ E-mail boxes is not stored in a way that allows reproducing the original E-mail communication. A potential attacker is not able to access the content of the participants’ E-mails.

Viewers and participants should gain insights into the tracked communication behavior by observing patterns in the E-mail communication. It is not intended to let viewers draw conclusions on individuals participating in the SCAD visualization project. The privacy issues are tackled by an appropriate software design as well as with technological means.
4. Description of the visualization

In this section the implementation of the E-mail visualization software is described. In Figure 1 an overview on the architecture of the system is given. The single parts of this architecture are illustrated in the following sections.

4.1. Aim of the Visualization

The implemented information system should allow participants as well as noninvolved people to grasp the shown communication network at a glance. Complicated means of interaction cannot be utilized in such a surrounding. In order to process feedback from users of the system, simple interaction is made possible by utilizing a webcam.

![Figure 1 Overview on the architecture](image)

4.2. Data acquisition

At the core of this information system is the acquisition of the E-mail communication. In order to address the potential privacy issues several technological means are applied during the data acquisition process. Every participant who wants their E-mail to show up in the SCAD visualization needs to opt-in on every E-mail they send.

An E-mail to be included in the visualization needs to be send to a dummy E-mail address in addition to the actual recipients of the E-mail.\(^1\) A customized dynamic IMAP crawler can now download information about the E-

---

\(^1\) This can be achieved by adding the dummy E-mail address to the CC field of each E-mail.
mail communication from this dummy E-mailbox. This IMAP crawler is a customized version of a crawler from the Condor software package, a dynamic social network analysis software formerly known as TeCFlow (Gloor & Zhao 2004). It is written in the Perl programming language. This data acquisition tool can download and process E-mail communication information from IMAP E-mailboxes. The tool extracts communication information from the E-mails and stores them into a MySQL database. The table schemes are designed to store information about a network. There is one table storing information about the nodes and one table that holds information about the edges between those nodes. The structure of the node table can be found in Figure 2, the structure of the edge table in Figure 3.

In order to allow a near real-time visualization of the occurring E-mail communication the IMAP crawler can be configured to frequently download information on new E-mails every N seconds from the dummy E-mailbox.

![Figure 2 Node table](image)

![Figure 3 Edge table](image)

The E-mail crawler is the first part of the introduced visualization system addressing the privacy issues emerging in E-mail analysis software. As soon as information on one E-mail is downloaded this E-mail is deleted from the server. Hereby all the E-mails sent to the dummy E-mail address are not stored together. A potential hijacker of the dummy E-mailbox is not able to access the E-mails of the participants. The IMAP crawler also only accesses the header of each E-mail, including the sender of the E-mail, its receiver, and the subject line, but not the content of the E-mail. The senders’ and the receivers’ E-mail addresses are ciphered before they are stored in the database of the visualization system. Also, the subject line of each E-mail is not stored together with the ciphered E-mail addresses to prevent linking of specific communication content to the (ciphered) E-mail addresses.

4.3. Visualizing the E-mail network

The part of the introduced information system most relevant and recognized by the user of this system is the visualization of the emerging communication networks.

Representing communication as a network with the communicating entities or actors shown as nodes and communication among them as the links between those nodes is a common visualization in social network analysis (Wasserman & Faust 1994). Because of its automatable collecting process E-mail communication is particularly well suitable for analyzing the resulting communication network.

The visualization of the network is implemented in the Java programming language. A visual arts framework built on top of Java, Processing, is utilized to implement the visualization. An enhancement to this framework, the

---

2 http://www.perl.org/
4 These nodes are the E-mail addresses of the participants in the visualization.
5 Each E-mails is stored as an edge.
6 http://www.java.com
Traer.Physics library\textsuperscript{9}, is used to calculate a three-dimensional particle system that determines the position of each node in the network. An OpenGL\textsuperscript{10} renderer visualizes those positions of nodes and the links among them. To allow a faster and visually more appealing visualization, a billboarding technique was developed to render the nodes of the communication network. Instead of assembling the spheres, representing the network’s nodes, of triangles an image (texture) is rendered at the nodes’ positions. Those images can have a much higher resolution than a sphere built from triangles. Since spheres look the same from each point of view they can be rendered by placing an image of a circle at the calculated position. Those images only need to be rotated to face the virtual camera. A screenshot of the resulting visualization is shown in Figure 4.

![Figure 4 A screenshot of the visualization](image)

4.4. Social feedback

Besides revealing the network structure of the participants’ E-mail communication the described software is also processing the content of the communication.\textsuperscript{11} The software tries to get a social feedback on the topics the participants discuss in their E-mails. In this experiment the authors want to explore how the world feels about the topics discussed in the collected E-mails. Querying the social media and micro-blogging service Twitter with topics extracted from currently sent E-mail subject lines generates this social feedback\textsuperscript{12}. Twitter allows its users posting short status messages, or tweets, on dedicated pages, their twitter feeds. The corresponding tweets are encoded into QR codes. Such a QR code, or quick response code, is a matrix code or two-dimensional bar code.\textsuperscript{13} The QR code generated from the received tweet is shown in the upper right corner of the visualization (cf. Figure 4). The QR code can be decoded by scanning it with the camera of a cell phone, either with built-in software or with additional software like the Kaywa reader\textsuperscript{14}.

\textsuperscript{7} http://processing.org

\textsuperscript{8} http://www.cs.princeton.edu/~traer/physics

\textsuperscript{9} OpenGL is a programming interface to produce graphics supported by several hardware manufacturers, http://www.opengl.org.

\textsuperscript{10} To be more specific: The content of the subject line of the E-mails, not the message content itself.

\textsuperscript{11} http://www.twitter.com

\textsuperscript{12} Specification of the QR code can be found at http://www.denso-wave.com/qrcode/qrstandard-e.html.

\textsuperscript{13} http://reader.kaywa.com
4.5. Interacting with the Visualization

The introduced software is built to be exhibited in a public space. Hence interacting with the software is not intended by common ways used with Personal Computers like keyboards or mice. In order to achieve a simple way for users to interact with the visualization a webcam and, again, QR codes are utilized. Simple commands can be passed to the visualization software by showing a command encoded via QR code to the webcam connected to visualization system. Those QR codes can be printed out in advance or be generated on the fly via a QR encoding software e.g. running on a cell phone. Participant of this E-mail visualization system can switch the visualization to display their very own local E-mail network by showing their QR encoded, ciphered E-mail address to the system’s webcam. The software then recognizes the E-mail address and only shows those E-mails in the visualization, which are sent or received by the given ciphered E-mail address.

5. Outlook

In this article an information system is described meant to visualize the E-mail communication of students and staff at a University. The evolvement of the E-mail network over time and the dynamic nature of this network is a main aspect to be shown by the described visualization software.

While gathering the E-mail data and processing it for the visualization, anonymized interaction data of the participants is stored in the system’s database as a by-product of the described process. In future studies these data can be analyzed with the means of dynamic social network analysis (Gloor 2005).

References